

This item is the archived preprint of:

Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci

Reference:

Betz Regina C., Petukhova Lynn, Ripke Stephan, Blaumeiser Bettina, Nöthen Markus, et al..- *Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci*

Nature communications - ISSN 2041-1723 - 6(2015), p. 1-8

DOI: <http://dx.doi.org/doi:10.1038/ncomms6966>

ARTICLE

Received 20 Jun 2014 | Accepted 25 Nov 2014 | Published xx xxx 2014

DOI: 10.1038/ncomms6966

Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci

Regina C. Betz^{1,*}, Lynn Petukhova^{2,3,*}, Stephan Ripke^{4,5,*}, Hailiang Huang^{4,5}, Androniki Menelaou⁶, Silke Redler¹, Tim Becker^{7,8}, Stefanie Heilmann^{1,9}, Tarek Yamany², Madeliene Duvic¹⁰, Maria Hordinsky¹¹, David Norris¹², Vera H. Price¹³, Julian Mackay-Wiggan², Annemieke de Jong², Gina M. DeStefano¹⁴, Susanne Moebus¹⁵, Markus Böhm¹⁶, Ulrike Blume-Peytavi¹⁷, Hans Wolff¹⁸, Gerhard Lutz¹⁹, Roland Kruse²⁰, Li Bian², Christopher I. Amos²¹, Annette Lee²², Peter K. Gregersen²², Bettina Blaumeiser²³, David Altshuler^{4,5}, Raphael Clynes^{2,24}, Paul I.W. de Bakker^{6,25}, Markus M. Nöthen^{1,9}, Mark J. Daly^{4,5} & Angela M. Christiano^{2,14}

Alopecia areata (AA) is a prevalent autoimmune disease with 10 known susceptibility loci. Here we perform the first meta-analysis in AA by combining data from two genome-wide association studies, and replication with supplemented ImmunoChip data for a total of 3,253 cases and 7,543 controls. The strongest region of association is the major histocompatibility complex, where we fine-map four independent effects, all implicating human leukocyte antigen-DR as a key aetiologic driver. Outside the major histocompatibility complex, we identify two novel loci that exceed statistical significance, containing ACOXL/BCL2L11(BIM) (2q13); GARP (LRRC32) (11q13.5), as well as a third nominally significant region SH2B3(LNK)/ATXN2 (12q24.12). Candidate susceptibility gene expression analysis in these regions demonstrates expression in relevant immune cells and the hair follicle. We integrate our results with data from seven other autoimmune diseases and provide insight into the alignment of AA within these disorders. Our findings uncover new molecular pathways disrupted in AA, including autophagy/apoptosis, transforming growth factor beta/Tregs and JAK kinase signalling, and support the causal role of aberrant immune processes in AA.

Q1

¹Institute of Human Genetics, University of Bonn, Bonn, Germany. ²Department of Dermatology, Columbia University, New York, New York 10032, USA. ³Department of Epidemiology, Columbia University, New York, New York, USA. ⁴Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. ⁵Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ⁶Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. ⁷German Center for Neurodegenerative Diseases, Bonn, Germany. ⁸Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany. ⁹Department of Genomics, Life and Brain Center, University Bonn, Bonn, Germany. ¹⁰Department of Dermatology, MD Anderson Cancer Center, Houston, Texas, USA. ¹¹Department of Dermatology, University of Minnesota, Minneapolis, Minnesota, USA. ¹²Department of Dermatology, University of Colorado, Denver, Colorado, USA. ¹³Department of Dermatology, University of California, San Francisco, San Francisco, California, USA. ¹⁴Department of Genetics and Development, Columbia University, New York, New York, USA. ¹⁵Institute of Medical Informatics, Biometry, and Epidemiology, University Duisburg-Essen, Essen, Germany. ¹⁶Department of Dermatology, University of Münster, Münster, Germany. ¹⁷Department of Dermatology and Allergy, Clinical Research Center for Hair and Skin Science, Charité-Universitätsmedizin Berlin, Berlin, Germany. ¹⁸Department of Dermatology, University of Munich, Munich, Germany. ¹⁹Dermatological Practice, Hair and Nail, Wesseling, Germany. ²⁰Dermatological Practice, Paderborn, Germany. ²¹Community and Family Medicine and Genetics, Dartmouth College, Hanover, New Hampshire, USA. ²²The Feinstein Institute for Medical Research, Manhasset, New York, USA. ²³Department of Medical Genetics, University of Antwerp, Antwerp, Belgium. ²⁴Department of Medicine, Columbia University, New York, New York, USA. ²⁵Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands. * These authors contributed equally to this work.

Q2

Correspondence and requests for materials should be addressed to A.M.C. (email: amc65@columbia.edu).

Q3

Alopecia areata (AA) is one of the most prevalent autoimmune diseases, with a lifetime risk of 1.7% (ref. 1), and is the most common cause of hair loss in children. In AA, aberrant immune destruction is targeted to the hair follicle, resulting in non-scarring hair loss that typically begins as patches, which can increase in size and coalesce and may progress to cover the entire scalp (alopecia totalis), and body as well (alopecia universalis). Disease prognosis is unpredictable and highly variable. Its aetiologic basis has remained largely undefined, creating barriers to the development of effective therapeutic strategies and an enormous unmet medical need^{2,3}.

Our first genome-wide association studies (GWAS) in AA identified associations in eight regions of the genome, which were subsequently confirmed in independent candidate gene studies^{4–7}. Associated loci outside the human leukocyte antigen (HLA) highlight particular immune response pathways and also implicate genes expressed in the hair follicle. For example, several regions contain genes with Treg functions, including *IL2RA*, *IL2/IL21*, *CTLA4* and *Eos*, whereas *ULBP3/ULBP6* implicate NKG2D-mediated cytotoxic T cells. Within the hair follicle, expression of *STX17* suggests a role for end-organ autophagy, while *PRDX5* implicates oxidative stress. A combined analysis of this GWAS and a subsequent replication study led to the identification of *IL13/IL4* and *KIAA0350/CLEC16A* as new gene loci⁵.

Q4

Here we perform a meta-analysis to expand our sample size, and identify two new loci that exceed our threshold for genome-wide significance and a third locus that is nominally significant. We identify transcripts and/or protein for candidate genes at all three loci in disease-relevant tissues. We perform imputation and fine-mapping of the HLA, identifying four independent associations that implicate HLA-DRβ1. Finally, CPMA of our data with published results from seven other autoimmune diseases identify molecular pathways shared by AA and one or more other disorders.

Q5

Results

In this study, we have increased our cohort size and performed a combined analysis of two GWAS using Illumina Human660W and Omni1-Quad BeadChips, analysing a total of 2,489 cases and 5,287 controls ascertained in the US and Central Europe (Supplementary Table 1). Association analyses are performed with logistic regression. In a meta-analysis of these data, nine of the previously implicated regions exceeded statistical significance ($P < 5 \times 10^{-8}$), with *STX17* achieving nominal significance ($rs10124366$; $P = 1.09 \times 10^{-5}$) (Fig. 1; Supplementary Data 1).

First, to resolve the major histocompatibility complex (MHC) association signal ($P = 4.91 \times 10^{-58}$ for the best single-nucleotide polymorphism (SNP), $rs9275516$), we used a published imputation and analysis protocol to perform fine-mapping (Supplementary Data 2)⁸. Conditional analysis revealed four independent variants located at the classical *HLA-DRA* and *HLA-DRB1* genes. The most significant variant was amino-acid position 37 in HLA-DRβ1 (omnibus P value = 4.99×10^{-73}). Of the five possible amino acids at this position, Leu (odds ratio (OR) = 1.56), Tyr (OR = 1.54) and Phe (OR = 1.19) conferred a higher risk of AA, whereas the other residues conferred lower risk (OR for Asn = 0.42; OR for Ser = 0.74). Adjusting for the effects of HLA-DRβ1 amino-acid position 37, we found an independent association due to an intronic SNP of *HLA-DRA*, $rs9268657$ (OR = 0.63, P value = 4.48×10^{-41}). Functional annotation of this SNP and its close proxies ($R^2 > 0.9$) reveal that they influence expression levels of *HLA-DRB1* (ref. 9). Adjusting for both HLA-DRβ1 amino-acid position 37 and $rs9268657$, we identified another independent association for the classical

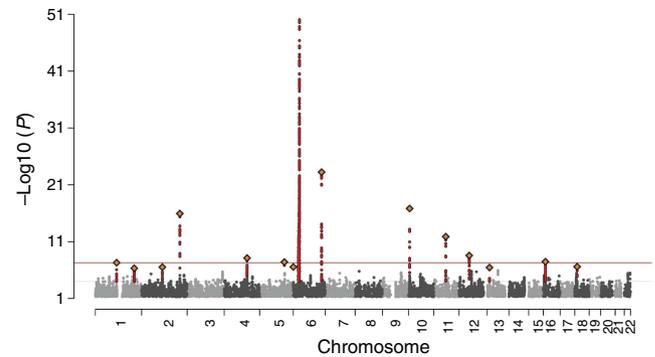


Figure 1 | Manhattan plot for genome-wide tests of association in meta-analysis. To conduct a meta-analysis across two GWAS, genotypes were imputed for each data set yielding 1.2 million SNPs. Standard association analysis with logistic regression including PC covariates was performed within each cohort and results were combined with s.e. weighted meta-analysis.

allele HLA-DRB1*04:01 (OR = 1.64, P value = 1.76×10^{-12}), confirming previous associations in candidate gene studies^{10,11}. Finally, adjusting for HLA-DRβ1 amino-acid position 37, $rs9268657$ and HLA-DRB1*04:01, amino-acid position 13 in HLA-DRβ1 was also significant (P value = 4.57×10^{-16}). Of the six possible alleles at this position, three increased risk (Tyr, OR = 1.41; Ser, OR = 1.35; Phe, OR = 1.09), two are protective (His, OR = 0.57; Gly, OR = 0.50) and one demonstrated no effect (Arg, OR = 0.98) (Supplementary Table 2). Further rounds of conditional analyses yielded no additional significant results ($P > 2.11 \times 10^{-6}$).

Collectively, these four independent associations in the MHC implicate HLA-DR as the primary risk factor in AA, presumably through antigen presentation, similar to other immune-mediated diseases¹². For example, HLA-DRβ1 amino-acid positions 13 and 37 contribute to P4 and P9 peptide-binding pockets, respectively; the disease associations of HLA-DRB1*04:01 are thought to be driven by the shared epitope (amino-acid residues 70–74)¹², which also occur within the peptide-binding cleft (Fig. 2). Polymorphic residues within peptide-binding pockets of HLA-DR influence binding affinities of peptides, and thus may shape the repertoire of autoantigens capable of triggering or perpetuating disease.

Next, we performed replication of SNPs in functionally relevant loci that achieved suggestive evidence for association in the meta-analysis ($5 \times 10^{-8} > P > 1 \times 10^{-5}$), utilizing an independent cohort of Central European ancestry. The Immunochip was used to genotype 318 cases and 1,688 controls and a Sequenom assay was used to genotype 85 SNPs not included on the Immunochip in a cohort of 764 cases and 568 controls. This analysis identified statistically significant associations in two novel genomic regions: chromosome 2q13 containing *ACOXL* and *BCL2L11* ($rs3789129$, $P = 1.51 \times 10^{-8}$, $OR_A = 1.3$), and chromosome 11q13.5 containing *C11orf30* and *LRCC32* ($rs2155219$, $P = 1.25 \times 10^{-8}$, $OR_T = 1.2$) (Table 1; Supplementary Data 3).

The association signal at chromosome 2q13 is located within an intron of *ACOXL*, and the region of association extends to include *BCL2L11* (Fig. 3). This region has been implicated in GWAS for two other autoimmune diseases: immunoglobulin-A nephropathy and primary sclerosing cholangitis¹³. *ACOXL* belongs to the acyl-coenzyme A oxidase gene family. While other family members play well-studied roles in peroxisomal beta-oxidation, very little is known about the function of this

gene. BCL-2-like 11, also known as BIM, is a member of the BCL-2 protein family and contains a Bcl-2 homology domain 3 (BH3) that interacts with other members to act as a pro-apoptotic factor. BIM has been widely studied in the context of apoptosis of immune cells, and has also been implicated in apoptosis of melanocytes¹⁴. More recently, data have emerged that implicate BIM in the destruction of some end-organs in autoimmune diseases, such as pancreatic beta cells in type 1 diabetes¹⁵. Furthermore, BIM has also been shown to regulate autophagy¹⁶, adding to growing evidence for the importance of this process in AA as recently shown by functional studies of STX17 (refs 17,18).

BIM is widely expressed in multiple cell types, including immune-related, epithelial and hair follicle cells¹⁹. We first

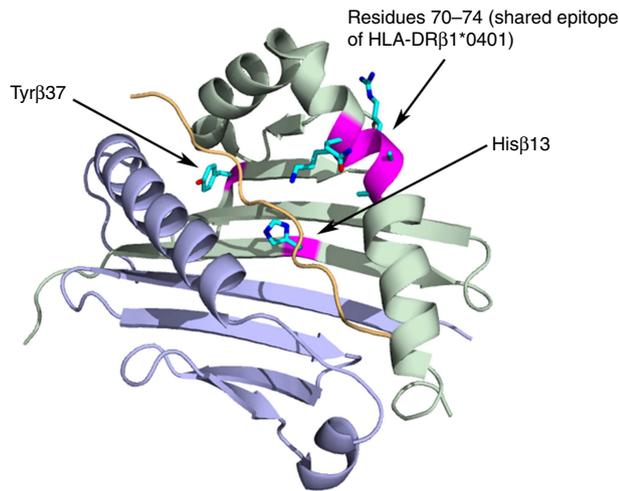


Figure 2 | Positions of amino-acid residues demonstrating independent association with alopecia areata. Structure of disease-associated HLA-DRB1*04*01 allele and polymorphic residues involved in susceptibility to AA. The peptide-binding cleft of an HLA-DR molecule is shown in cartoon representation with the α -chain coloured in blue and β -chain in green. The MHC-bound peptide is shown in cartoon representation and coloured yellow. Residues His13 β and Tyr37 β corresponding to amino-acid positions with independent disease association are shown in a stick model along with the shared epitope (residues 70–74). Crystal structure of HLA-DR4 bound to melanocytes lineage-specific antigen gp120 was used (PDB 4IS6).

performed reverse transcription (RT)–PCR analysis of *BIM* expression in immune cells and scalp hair follicles (SHFs). We detected *BIM* transcript in whole-peripheral blood mononuclear cells (PBMCs), including T, natural killer and B cells, and monocytes (MCs) (Fig. 4a). In plucked human SHFs, we observed expression of an alternative splice isoform, *BIM-S*, the most pro-apoptotic variant (Fig. 4a)^{20,21}. To characterize protein localization within hair follicles, we performed immunofluorescence staining on both human and mouse hair follicles. In the human hair follicle, BIM is highly expressed in the bulb of the hair follicle, especially within the matrix cells, in a pattern strikingly restricted to early catagen, but not anagen or telogen (Fig. 4b). The hair bulb matrix is the principal location of differentiated, pigment-producing melanocytes, which are postulated to undergo apoptosis during catagen²². In mouse hair follicles, BIM localized to the apoptosing strand and lower portion of the catagen hair follicle, consistent with its expression in human hairs. Remarkably, BIM expression was restricted to catagen hair follicles, where immunofluorescence staining performed on anagen or telogen hairs produced no signal (Fig. 4c, upper panel). Finally, to investigate a possible role of BIM in AA pathogenesis, we used a mouse model that recapitulates genetic and molecular profiles of human AA²³. C3H/HeJ spontaneously develops AA, and we performed immunofluorescence staining on affected and unaffected skin. We observed a striking and widespread increase in BIM expression throughout affected skin and hair follicles, which was not restricted to catagen hair follicles and appeared further upregulated in lesional skin (Fig. 4c, lower panel). The regression stage of a normal cycling hair follicle, catagen, is characterized by apoptosis and cell death. We postulate that dysregulation of BIM in hair-matrix keratinocytes and/or hair follicle melanocytes contributes to the early entry into dystrophic catagen in AA hair follicles in active disease.

The second novel region that exceeded statistical significance at chromosome 11q13.5 (rs2155219, $P = 1.25 \times 10^{-8}$) contains three gene transcripts, including *C11orf30*, *GUCY2E* and *GARP* (*LRRC32*). GWAS have implicated this region in several autoimmune and inflammatory diseases (www.genome.gov/gwas-studies). *GARP* is expressed on activated Tregs and binds and augments transforming growth factor beta bioavailability^{23–26}, functioning to induce *FOXP3* expression and Treg differentiation. Expression of *GARP* on activated human Treg cells correlates with

Table 1 | Candidate genes in AA GWAS regions.

Locus	Genes of interest	SNP	Chr	BP	A1A2	P	OR
6p21.32	HLA-DQB1	rs9275524	6	32,783,087	TC	1.8E – 60	0.52
10p15.1	IL15RA,IL2RA	rs3118470	10	6,141,719	TC	7.7E – 21	0.71
2q33.2	CD28,CTLA4,ICOS	rs231775	2	204,440,959	AG	2.2E – 20	0.72
6q25.1	RAET1L,ULBP3	rs12183587	6	150,396,301	TG	5.9E – 24	1.48
11q13	PRDX5	rs574087	11	63,859,524	AG	8.7E – 14	1.32
12q13	IKZF4 (Eos),ERBB3	rs2292239	12	54,768,447	TG	4.4E – 09	1.25
4q27	IL21,IL2	rs7682481	4	123,743,476	CG	4.8E – 09	1.23
5q31.1	IL13,IL4	rs848	5	132,024,399	AC	4.8E – 09	1.27
2q13	ACOXL,BCL2L1(BIM)	rs3789129	2	111,414,511	AC	1.5E – 08	1.31
11q13.5	GARP(LRRC32)	rs2155219	11	75,976,842	TG	4.1E – 08	1.21
1p13.2	PTPN22	rs2476601	1	114,179,091	AG	8.9E – 08	1.34
12q24.12	SH2B3(LNK),ATXN2	rs653178	12	110,492,139	TC	1.6E – 07	0.84
16p13.13	CIITA,CLEC16A,SOCS1	rs3862469	16	11101581	TC	1.7E – 07	0.82
9q31.1	STX17, NR4A3	rs10124366	9	101,727,524	AG	1.1E – 05	0.83

GWAS, genome-wide association studies; OR, odds ratio; SNP, single-nucleotide polymorphism.

Fourteen regions in the genome have demonstrated association with AA by GWAS and replication. For each region, genes of interest are listed, along with the most significant SNP, its location in the genome, P value and OR obtained with logistic regression from the combined analysis ($N = 10,796$). P values and effect estimates from each analytic stage are in Supplementary Table 3.

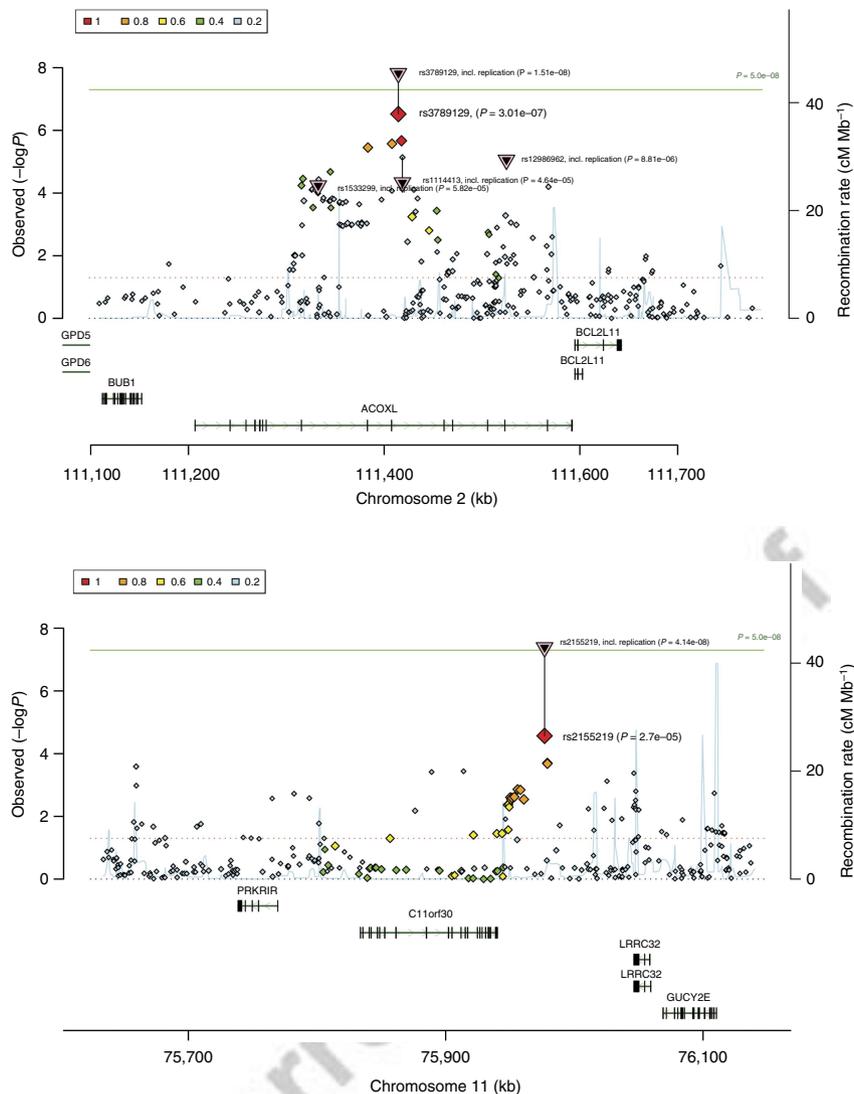


Figure 3 | Detailed map of associated SNPs and gene locations for newly identified loci. Two regions in the genome exceeded statistical significance when the replication data were combined with the meta-analysis results ($N = 10,796$) and analysed with logistic regression. **(a)** Chromosome 2q13 includes ACOXL and BCL2L11. **(b)** Chromosome 12q24.12 included C11orf30 and LRRC32 (GARP).

their increased suppressive activity, and GARP knockdown reduces their suppressive activity^{24,27}. Not surprisingly, we found that GARP is highly expressed in whole PBMCs, consistent with previous reports. Unexpectedly, we also detected a strong signal for GARP in plucked control SHFs, a site not previously described and perhaps atypical for the expression of a Treg protein (Fig. 4). This suggests a potential role for GARP in hair biology and implicates significance of its disruption in AA at the site of pathology for the phenotype.

A third novel region on chromosome 12q24.12 that achieved suggestive evidence for association ($P = 1.3 \times 10^{-7}$) is of interest, because several genes harboured within the region functionally align with other associated genes. This region has demonstrated associations with multiple autoimmune diseases, and contains 10 genes including SH2B adaptor protein 3 (*SH2B3*) and aldehyde dehydrogenase 2 family (*ALDH2*). *SH2B3*, also known as LNK, is a key negative regulator of cytokine signalling via receptor tyrosine kinases and JAK signalling. Two missense variants that confer increased cytokine production and enhanced signalling have been reported in the literature (rs3184504 and rs72650673)^{28,29}. We genotyped these two functional

polymorphisms in a sample of 96 chromosomes from AA patients carrying the AA-associated risk allele (rs653178*G), and found that 93 chromosomes also carried the rs3184504*T risk allele ($f = 0.97$). This increase in allele frequency over what would be expected among a sample of European Americans ($f = 0.51$ in Exome Variant Server ESP6500) indicates that there is strong LD between the AA-associated risk variant and this functional polymorphism, and suggests that R262W could be contributing to AA pathogenesis. The rs72650673*A allele was not found in our sample, consistent with its low allele frequency among European Americans ($f = 0.002$ in Exome Variant Server ESP6500). We found by RT-PCR analysis that LNK is highly expressed in whole PBMCs, as well as T cells, natural killer, B cells and MCs, but did not detect expression in plucked SHFs (Fig. 4). *ALDH2* has been identified as a citrullinated antigen in RA patients³⁰. This gene has been studied in the skin, with expression localizing to epidermis, sebaceous glands and hair follicles, where it is hypothesized to reduce the accumulation of oxidative stress-induced aldehydes³¹.

The strong association with HLA class II in AA points towards the involvement of CD4+ T cells in the pathogenesis.

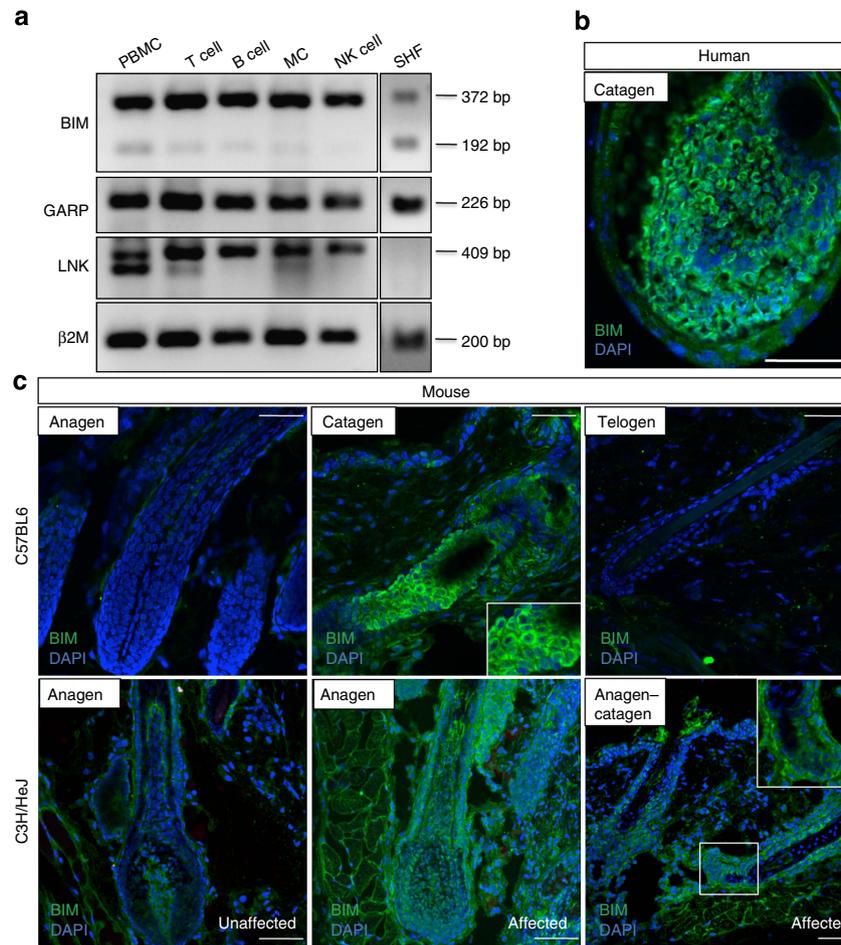


Figure 4 | Characterization BIM, GARP and LNK expression. (a) *BIM*, *GARP* and *LNK* expression in human immune cells and hair follicles. RT-PCR was performed on *BIM*, *GARP* and *LNK* to determine gene expression in T cells, natural killer cells (NK), B cells, monocytes (MC), peripheral blood mononuclear cells (PBMCs) and scalp hair follicles (SHF). β 2M PCR was used as a loading control for each complementary DNA. Two splice variants are observed for *BIM* expression: *BIM-S* (192 bp) and *BIM-L* (372 bp). Expected amplicon sizes: *BIM* (372 bp and 192 bp), *GARP* (226 bp) and *LNK* (409 bp). (b) Immunofluorescence staining in human hair follicles reveals that *BIM* is highly expressed in matrix cells of the catagen hair bulb. (c) In healthy hair follicles from the C57BL/6 mouse strain, *BIM* is strongly expressed in the apoptosing strand of mouse catagen hair follicles and is absent from anagen and telogen hair follicles. Immunofluorescence staining on affected and unaffected skin from the AA mouse model strain, C3H/HeJ, reveals that *BIM* expression levels and the localization pattern are aberrant in affected hair follicles compared with unaffected C3H/HeJ and C57BL/6 hair follicles. DAPI, 4',6-diamidino-2-phenylindole.

The importance of this T-cell subset is supported by the presence of perifollicular CD4⁺ T cells in addition to intrabulbar CD8⁺ T cells in human AA. Finally, we analysed these data in the context of other autoimmune diseases by performing a cross-phenotype meta-analysis using the 107 SNPs used in the original description of the analysis and data from seven other autoimmune diseases³². We found 50 SNPs that had significance at or near significance across two or more diseases, and clustering of these according to their association with disease identified five groups (Fig. 5a,b). Protein–protein interaction graphs of each SNP cluster demonstrate that for several of the groups, the proteins coded from these regions interact either directly or via an intermediary to a significant degree (Fig. 5c). Integration of AA into the CPMA demonstrates mechanistic alignment with coeliac disease, type 1 diabetes, rheumatoid arthritis, multiple sclerosis and Crohn's disease, with most of the overlap coming from the first cluster of genes.

Collectively, this meta-analysis for the first time resolves HLA associations and demonstrates the pivotal aetiological role of HLA-DR. Furthermore, the identification of specific residues in HLA-DR β 1 that are over-represented among AA patients will

allow us to better model peptide class II MHC interactions and thus predict autoantigens in AA. Associations outside HLA provide new evidence for the importance of Treg maintenance and immune response pathways in AA. Furthermore, autophagy and apoptosis are emerging as processes of aetiological importance in AA. These insights will allow us to better understand the molecular taxonomy of autoimmune diseases and the alignment of AA within this class of disorders. Importantly, as GWAS help to resolve disease mechanisms and identify pathogenic pathways perturbed in AA and autoimmunity in general, these approaches advance the field towards precision medicine in autoimmunity.

Methods

Patient population. All participating studies were reviewed and approved by the Institutional Review Boards and ethics committees at Columbia University; MD Anderson Cancer Center; University of Minnesota, Minneapolis; University of Colorado, Denver; University of California, San Francisco; and the Universities of Bonn, Düsseldorf, Münster, Berlin, Hamburg and Munich, Germany; and Antwerp, Belgium, and were conducted in accordance with the Declaration of Helsinki Principles. All study subjects provided written informed consent.

For the previously published GWAS and the ImmunoChip samples, cases were ascertained through the National Alopecia Areata Registry, which recruits patients

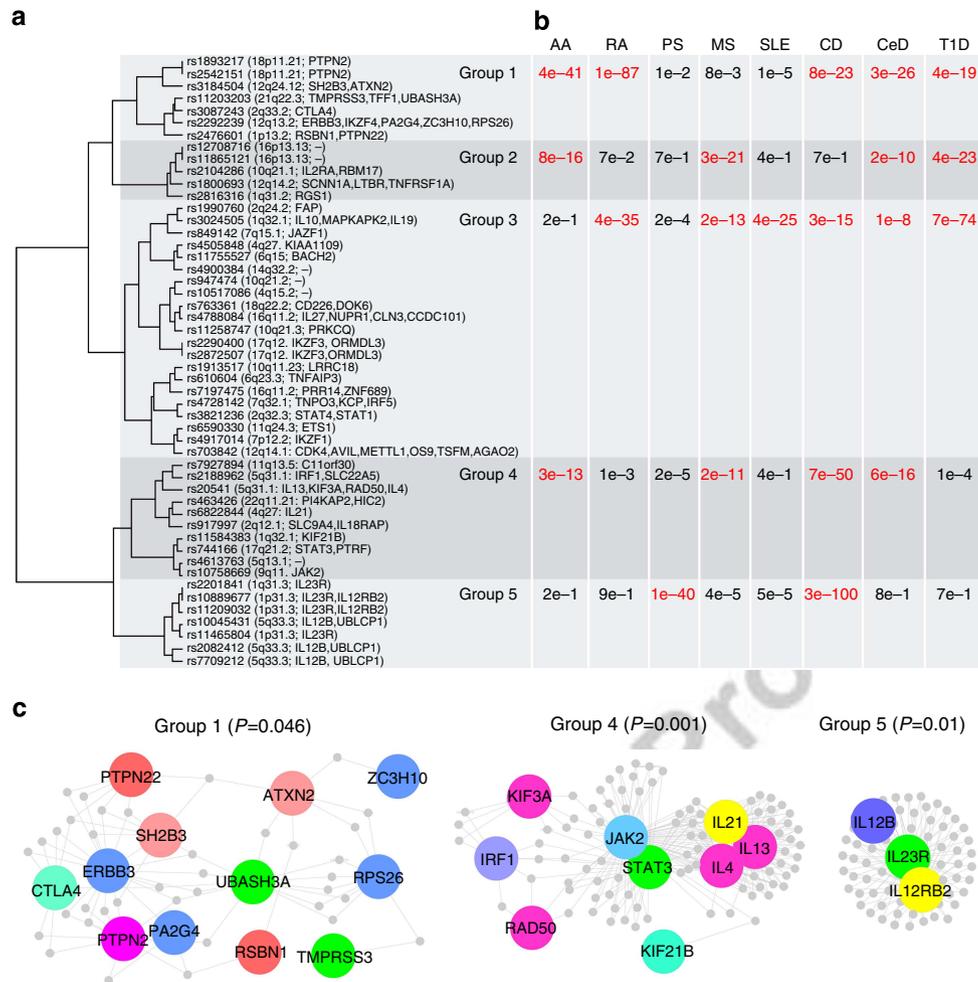


Figure 5 | Cross-phenotype meta-analysis revealing functional clusters of genes and proteins indicating patterns of shared disease mechanisms across autoimmune diseases. (a) Fifty SNPs with evidence of association to more than one autoimmune disease ($P_{\text{cpma}} < 0.01$) are clustered by association with disease. (b) For each cluster and disease pair, a cumulative association statistic was calculated using Fisher’s omnibus test to combine P values. The varying pattern of disease association for each cluster suggests each group represents a distinct co-morbid mechanism. (c) Proteins coded within a 100-Mb window centred on each SNP within each cluster are depicted in protein–protein interaction maps. Three of the five clusters have significant protein inter-connectivity (permuted $P < 0.05$).

in the US through five clinical sites and confirms diagnosis. Control data for these samples were obtained from publically available data. There were two sources of control data for the GWAS⁴. First, a data set was obtained from subjects enrolled in the New York Cancer Project³³ and genotyped as part of previous studies³⁴. Second, a data set was obtained from the CGEMS breast³⁵ and prostate³⁶ cancer studies (<http://cgems.cancer.gov/data/>). The ImmunoChip control data were obtained from the NIDDK inflammatory bowel disease genetics consortium (<http://medicine.yale.edu/intmed/ibdgc/index.aspx>). The cases and controls for the previously unreported GWAS were ascertained from outpatient clinics, private dermatology practices and via AA self-support groups in Belgium and Germany. Inclusion criteria followed published guidelines³⁷, and additionally included diagnosis by a trained and experienced clinician. European GWAS controls included population-based controls established within the National Genome Research Network for use as universal controls (PopGen³⁸, KORA³⁹), and the Heinz Nixdorf Recall Study⁴⁰ and an additional population-based cohort from the Heinz Nixdorf Recall⁴⁰ study. Sample counts by ascertainment location and genotyping platform are summarized in Supplementary Table 1.

Genotyping. DNA was extracted from peripheral blood leukocytes by salting out with saturated NaCl solution according to standard methods, or by using a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany), in accordance with the manufacturer’s instructions. Whole-genome genotyping for the meta-analysis was performed on either Illumina HumanHap550 BeadChip or the Illumina Omni express, as detailed in Supplementary Table 1. The replication cohort was genotyped with the Illumina ImmunoChip. In addition, SNPs selected for replication, which were not present on the ImmunoChip, were genotyped on the MassArray system using a Sequenom Compact MALDI-TOF device and iPLEX

Gold reagents (Sequenom, San Diego, CA) in multiplex reactions. Primer sequences and standard assay conditions are available upon request. Standard quality-control (QC) metrics were employed, removing data for SNPs and samples with call rates less than 90. Additional genotyping was performed by standard PCR-based techniques.

Discovery QC and imputation. Technical QC was performed with QC conducted on each data set separately using a common approach. The following QC parameters were applied: (i) missing rate per SNP < 0.05 (before sample removal below), (ii) missing rate per individual < 0.02 , (iii) heterozygosity per individual ± 0.2 , (iv) missing rate per SNP < 0.02 (after sample removal above), (v) missing rate per SNP difference in cases and controls < 0.02 , (vi) Hardy–Weinberg equilibrium (in controls) $P < 10^{-6}$, (vii) Hardy–Weinberg equilibrium (in cases) $P < 10^{-10}$. Study sample sizes varied between 1,200 and 3,000 individuals. The number of SNPs per study after QC varied between 420,000 and 640,000. On average, the QC processes excluded 44 individuals per study (with a range of 13–107 individuals) and 13,000 SNPs per study (with a range of 3,000–20,000 SNPs). After QC, the GWAS data sets together comprised 2,489 cases and 5,287 controls and, for the next steps of the ‘genetic QC’ analysis, a set of 221,784 SNPs common to all platforms and successfully genotyped in each GWAS sample was extracted. These SNPs were then further pruned to remove LD (leaving no pairs with $r^2 > 0.05$) and lower-frequency SNPs (minor allele frequency < 0.05), leaving 64,099 SNPs suitable for robust relatedness testing and population structure analysis (see below). Imputation of untyped SNPs was performed within each study in batches of 300 individuals. These batches were randomly drawn to keep the same case-control ratio as in the total sample from that study. We used Beagle 3.13. Imputation was performed with CEU + TSI HapMap phase 3 data (UCSC hg18/

Q9

Q10

Q11

NCBI 36) using a chunk size of 10 Mb with 410 phased haplotypes comprising 1,233,578 SNPs, using default parameters. Lambda-GC was carefully monitored before and after imputation.

Genetic QC included relatedness testing and principal component analyses based on 64,099 LD-independent SNPs, present on all platforms in this study. Relatedness testing was done with PLINK⁴¹, reporting pairs with genome identity (pi-hat) > 0.9 as 'identical samples' and with pi-hat > 0.2 as being closely related. After random shuffling, one individual from each related pair was excluded from downstream analysis. From groups with multiple related pairs (for example, a family), only one individual was kept.

Q12 **Discovery principal component analysis.** Principal component estimation was done with the same collection of SNPs on the non-related subset of individuals⁴². We estimated the first 20 principal components and tested each of them for phenotype association (using logistic regression with study indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using Lambda-GC (the genomic control inflation factor based on the median χ^2) after genome-wide association of the specified principal component. On the basis of this, we decided to include principal components 1, 2, 3, 4, 5 and 6 for downstream analysis as associated covariates.

Association analysis. The primary scan had a total of 2,332 QC+ cases and 5,233 QC+ controls. Association testing was carried out in PLINK⁴¹, using the dosage data from the imputation and using the first six principal components as covariates, chosen as described above from the first 20 principal components. The four data sets had Lambda-GC values of 1.025, 1.041, 1.040 and 1.042.

A fixed-effect meta-analysis was performed using ORs and s.e. The final meta-analysis had a Lambda-GC of 1.076.

MHC imputation and association analysis. Starting from the genotyped SNPs, we imputed classical HLA alleles and corresponding polymorphic amino acids within classical HLA proteins, following a recently described protocol⁶. In total, 8,271 variants including SNPs, amino acids and classical HLA alleles at 2- and 4-digit resolution were tested for association in the MHC region. Logistic regression was applied for testing for association, including the six first principal components and a dummy variable for the cohort id of each sample. Single-test association was performed for all types of variants, as well as omnibus tests for the amino acids. The conditional analyses were performed using the initial PCA covariates, the dummy variable for the cohort, and adjusting for the top and conditional hits.

Q13 **Q14** **Analysis of immunochip data.** To identify the sample ethnicities in the case cohort, we calculated the principal components (PC) 1 and 2 using Hapmap 3, and mapped the case samples to the principal components. Samples that do not cluster with the Hapmap3 European samples were discarded. The NIDDK control cohort has been cleaned for sample ethnicity and is thus not necessary to undergo such procedure. For both case and control cohort, we then removed samples that have missing genotype rate > 2% and variants that have missing genotype rate > 2% for each cohort. For the control cohort, we additionally removed variants that fail the Hardy-Weinberg equilibrium test (P value $< 1 \times 10^{-5}$). The case and the control cohorts were then merged, and we removed variants that have different missingness in cases and in controls (P value $< 1 \times 10^{-5}$) and samples that have high heterozygosity (> 3 s.d.). We used PLINK to infer the genetic relationships between the remaining samples and we randomly removed one of the samples for each related sample pair (PI_HAT > 0.5)⁴¹. The final cleaned data set has 318 cases and 1,688 controls. We tested the association between the genetic variants and AA using PLINK with the principal components 1–5 as covariates.

Q15 **RT-PCR and immunofluorescence staining on human and mouse tissues.** RT-PCR was performed on total RNA extracted from plucked human SHFs and T cells, natural killer cells, B cells, MCs, PBMCs that were isolated by FACS from whole blood. The primer sequences of the *BIM*, *GARP*, *LNK* and $\beta 2M$ genes are as follows: *BIM* F: 5'-TAAGTTCTGAGTGTGACCGAGA-3', R: 3'-CCATTGCAC TGAGATAGTGGTTG-5'; *GARP* F: 5'-CGCTCCGAGACTCATCTAC-3', R: 5'-AGGTGCTCAAGAAAGCTGTGCG-3'; *LNK* F: 5'-GTGGGAATACGTGCTCA CTT-3', R: 5'-TGTCCACGACCGAGGAAA-3'; $\beta 2M$ F: 5'-GAGGCTATCCAG CGTACTCCA-3', R: 5'-CGGCAGGCATACTCATCTTTT-3'.

Immunofluorescence staining was performed on human and mouse samples. Human skin and hair follicle sections were obtained from control occipital scalp biopsies discarded during surgery and considered to be non-human subject research under 45 CFR Part 46 (Institutional Review Board exempted). Mouse anagen skin (day 30), catagen skin (day 42) and telogen skin (day 50) was collected from the dorsal region and embedded in optimal cutting temperature for sectioning (10 μ m). Slides were fixed with 50% MeOH/50% acetone for 10 min at 220 μ C, washed with 1 \times phosphate-buffered saline and then blocked with 2% fish skin gelatin (Sigma Aldrich, MO, USA). The α -BIM antibody (Rabbit) (Biorbyt, orb10190) was used at a concentration of 1:200 for both mouse and human studies. Slides were washed, incubated with the Alexa Fluor 488 donkey anti-rabbit IgG (Molecular Probes, Invitrogen) secondary antibody (1:800 in

1 \times phosphate-buffered saline), mounted with VECTASHIELD mounting medium with DAPI (Vector Laboratories, Burlingame, CA, USA) and imaged using a LSM 5 laser-scanning Axio Observer Z1 confocal microscope (Carl Zeiss).

Cross-phenotype meta-analysis. The protocol for this analysis was adopted from previously described work on cross-phenotype meta-analysis³². Using Dataset S1 from the Supplementary Data of the original analysis, we added in AA association P values for the 107 SNPs originally investigated. For SNPs with no P value, we used Broad Institutes SNAP to find a proxy SNP with $r^2 > 0.9$ (<http://www.broadinstitute.org/mpg/snap/>). The CPMA statistic was then recalculated for all 107 SNPs using a code (found here: <http://coruscant.itmat.upenn.edu/software.html>) implementing the original description. All SNPs meeting a significance threshold of < 0.01 were included in clustering analysis. Clustering was done by first binning P values into four groups based on magnitude and clustered using the Cluster package. Cumulative association statistics were calculated in R using Fisher's omnibus test. Protein-protein interaction analysis was performed using DAPPLE, a publicly available web application (<http://www.broadinstitute.org/mpg/dapple>). Settings for DAPPLE included the 1000 genome assembly, 5,000 permutations and 50-kb regulatory region upstream and downstream. All statistical analyses were performed in R version 3.0.1.

References

- Safavi, K. H., Muller, S. A., Suman, V. J., Moshell, A. N. & Melton, 3rd L. J. Incidence of alopecia areata in Olmsted County, Minnesota, 1975 through 1989. *Mayo Clin. Proc.* **70**, 628–633 (1995).
- Delamere, F. M., Sladden, M. M., Dobbins, H. M. & Leonardi-Bee, J. Interventions for alopecia areata. *Cochrane Database Syst. Rev.* CD004413 (2008).
- Lebwohl, M. *Treatment of Skin Disease: Comprehensive Therapeutic Strategies* 2nd edn (Mosby/Elsevier, 2006).
- Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466**, 113–117 (2010).
- Jagielska, D. *et al.* Follow-up study of the first genome-wide association scan in alopecia areata: IL13 and KIAA0350 as susceptibility loci supported with genome-wide significance. *J. Invest. Dermatol.* **132**, 2192–2197 (2012).
- John, K. K. *et al.* Genetic variants in CTLA4 are strongly associated with alopecia areata. *J. Invest. Dermatol.* **131**, 1169–1172 (2011).
- Redler, S. *et al.* Investigation of selected cytokine genes suggests that IL2RA and the TNF/LTA locus are risk factors for severe alopecia areata. *Br. J. Dermatol.* **167**, 1360–1365 (2012).
- Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
- Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
- Entz, P. *et al.* Investigation of the HLA-DRB1 locus in alopecia areata. *Eur. J. Dermatol.* **16**, 363–367 (2006).
- Colombe, B. W., Lou, C. D. & Price, V. H. The genetic basis of alopecia areata: HLA associations with patchy alopecia areata versus alopecia totalis and alopecia universalis. *J. Invest. Dermatol. Symp. Proc.* **4**, 216–219 (1999).
- Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296 (2012).
- Melum, E. *et al.* Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat. Genet.* **43**, 17–19 (2010).
- Bouillet, P., Cory, S., Zhang, L. C., Strasser, A. & Adams, J. M. Degenerative disorders caused by Bcl-2 deficiency prevented by loss of its BH3-only antagonist Bim. *Dev. Cell* **1**, 645–653 (2001).
- Santin, I. *et al.* PTPN2, a candidate gene for type 1 diabetes, modulates pancreatic beta-cell apoptosis via regulation of the BH3-only protein Bim. *Diabetes* **60**, 3279–3288 (2011).
- Luo, S. Q. & Rubinsztein, D. C. BCL2L1/BIM: a novel molecular link between autophagy and apoptosis. *Autophagy* **9**, 104–105 (2013).
- Hamasaki, M. *et al.* Autophagosomes form at ER-mitochondria contact sites. *Nature* **495**, 389–393 (2013).
- Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151**, 1256–1269 (2012).
- O'Reilly, L. A. *et al.* The proapoptotic BH3-only protein bim is expressed in hematopoietic, epithelial, neuronal, and germ cells. *Am. J. Pathol.* **157**, 449–461 (2000).
- Nogueira, T. C. *et al.* GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. *PLoS Genet.* **9**, e1003532 (2013).
- Jiang, C. C. *et al.* Apoptosis of human melanoma cells induced by inhibition of B-RAFV600E involves preferential splicing of bimS. *Cell Death Dis.* **1**, e69 (2010).
- Tobin, D. J. Aging of the hair follicle pigmentation system. *Int. J. Trichol.* **1**, 83–93 (2009).

23. Battaglia, M. & Roncarolo, M. G. The Tregs' world according to GARP. *Eur. J. Immunol.* **39**, 3296–3300 (2009).
24. Tran, D. Q. *et al.* GARP (LRRC32) is essential for the surface expression of latent TGF-beta on platelets and activated FOXP3+ regulatory T cells. *Proc. Natl Acad. Sci. USA* **106**, 13445–13450 (2009).
25. Stockis, J., Colau, D., Coulie, P. G. & Lucas, S. Membrane protein GARP is a receptor for latent TGF-beta on the surface of activated human Treg. *Eur. J. Immunol.* **39**, 3315–3322 (2009).
26. Wang, R. *et al.* GARP regulates the bioavailability and activation of TGFbeta. *Mol. Biol. Cell* **23**, 1129–1139 (2012).
27. Wang, R. *et al.* Expression of GARP selectively identifies activated human FOXP3+ regulatory T cells. *Proc. Natl Acad. Sci. USA* **106**, 13439–13444 (2009).
28. Zhernakova, A. *et al.* Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am. J. Hum. Genet.* **86**, 970–977 (2010).
29. McMullin, M. F., Wu, C., Percy, M. J. & Tong, W. A nonsynonymous LNK polymorphism associated with idiopathic erythrocytosis. *Am. J. Hematol.* **86**, 962–964 (2011).
30. Wegner, N. *et al.* Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.* **233**, 34–54 (2010).
31. Cheung, C., Davies, N. G., Hoog, J. O., Hotchkiss, S. A. M. & Pease, C. K. S. Species variations in cutaneous alcohol dehydrogenases and aldehyde dehydrogenases may impact on toxicological assessments of alcohols and aldehydes. *Toxicology* **184**, 97–112 (2003).
32. Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254 (2011).
33. Mitchell, M. K., Gregersen, P. K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health* **81**, 301–310 (2004).
34. Plenge, R. M. *et al.* TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *New Engl. J. Med.* **357**, 1199–1209 (2007).
35. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
36. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
37. Olsen, E. A. *et al.* Alopecia areata investigational assessment guidelines—Part II. National Alopecia Areata Foundation. *J. Am. Acad. Dermatol.* **51**, 440–447 (2004).
38. Krawczak, M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **9**, 55–61 (2006).
39. Wichmann, H. E., Gieger, C., Illig, T. & Group, M. K. S. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**(Suppl 1): S26–S30 (2005).
40. Schermund, A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am. Heart J.* **144**, 212–218 (2002).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
42. DeStefano, G. M. *et al.* Mutations in the cholesterol transporter gene ABCA5 are associated with excessive hair overgrowth. *PLoS Genet.* **10**, e1004333 (2014).

Acknowledgements

We are deeply indebted to the many patients and their family members who participated in this study. We are thankful to Drs Lawrence Shapiro and Tatyana Gindin for guidance on the structural effects of HLA associations. We thank Spandon V. Shah, Holly Jiang, Matthew Ding and Steven Chiu for their assistance and contributions in isolating the immune cells by FACS and Jane E. Cerise for her bioinformatic expertise. We are most grateful for the support of the National Alopecia Areata Foundation for funding the initial studies, and Ms Vicki Kalabokos and her staff at NAAF for their efforts on our behalf. The US patient cohort was collected and maintained by the National Alopecia Areata Registry (N01-AR62279) (to M.D.). Some controls were drawn from the Heinz Nixdorf Recall Study cohort, which was established with the support of the Heinz Nixdorf Foundation. R.C.B. and M.M.N. are members, T.B. is an associate member, of the DFG-funded Excellence Cluster ImmunoSensation. R.C.B. is a recipient of a Heisenberg Professorship of the German Research Foundation (DFG). This work was supported in part by the DFG grant BE 2346/5-1, as well as by local funding (BONFOR) to R.C.B. and USPHS NIH/NIAMS grants R01AR52579 and R01AR56016 (to A.M.C.).

Author contributions

R.C.B., M.M.N., S.R., M.D., V.H.P., M.H., D.N. and J.M.-W. participated in phenotyping, diagnosis and access to patient samples. L.P. performed technical aspects in preparation of samples for genotyping and data transfer. H.H., L.P., S.R., T.Y., T.B., A.M. and P.I.W.d.B. performed statistical analyses. G.M.D. performed immunofluorescence staining and RT-PCR experiments. A.L. and P.K.G. provided control samples and performed genotyping, as well as insight into autoimmune diseases. C.I.A. provided additional statistical advice and control samples. L.P. contributed to study design, interpretation of results and drafting of the manuscript. R.C.B., M.M.N., M.D., D.A. and A.M.C. provided conceptual guidance to the project. A.d.J., R.C., A.M. and P.I.W.d.B. provided expertise on interpreting HLA results. A.M.C. additionally supplied input into the functional significance of candidate genes, supervision of lab personnel, management of collaborations, preparation of the manuscript and all reporting requirements for granting agencies.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Betz, R. C. *et al.* Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat. Commun.* **5**:5966 doi: 10.1038/ncomms6966 (2014).

Q19

Q17

Q18